

けいはんなR&Dフェア2020

WEB展示

～ニューラルネットを用いた高品質話速変換技術～

○ニューラルネットを用いた高品質話速変換技術

NICT先進的音声翻訳研究開発センター（ASTREC）先進的翻訳技術研究所

令和2年11月5日～7日の3日間、けいはんな学研都市に立地または関連する研究機関や自治体、企業、大学などによる、最先端技術の研究発表イベントが開催されました。新型コロナウイルス感染症の影響により特設WEBサイトによるバーチャル開催となりました。基調講演や技術講演のほか、研究成果をまとめたパネルや録画動画により活動を紹介するWEB展示が、合計49のテーマで行われました。



「ニューラルネットを用いた高品質話速変換技術」では、音声の品質を保ちながら、会話を速くしたり遅くしたりできる技術が紹介されました。近年のディープラーニングによって機械翻訳の精度は飛躍的に向上しましたが、このディープラーニング技術が脳の機能に類似したニューラルネットワーク（神経網）によるものであることから、こうした機械翻訳のモデルがニューラル機械翻訳と呼ばれています。

話速変換技術とは、音の高さを変えずに音声を伸縮する技術です。高齢者や難聴者は通常よりもゆっくりと話す音声の方が聞き取りやすく、視覚障がい者は高速再生された音声を好む傾向があります。そのため、それぞれの状況で話速変換技術が用いられてきましたが、音声をそのままゆっくり、または高速で再生してしまうと音の高さまで変わってしまいます。音の高さを変えずに、再生速度のみを変える従来方式の話速変換技術では、音質が劣化するという課題がありました

が、ニューラルネットを用いた話速変換技術により、高品質な音声を瞬時に合成することが可能になりました。この高品質話速変換は、音声の音響特徴量を分析し、時間方向に伸縮した特徴量を音声波形生成ニューラルネットワークに入力することで可能になります。

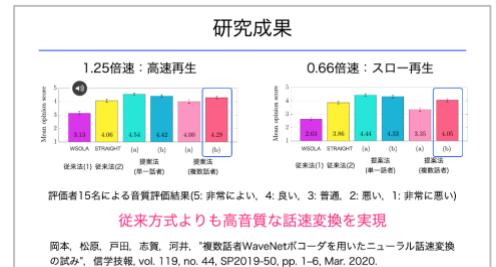
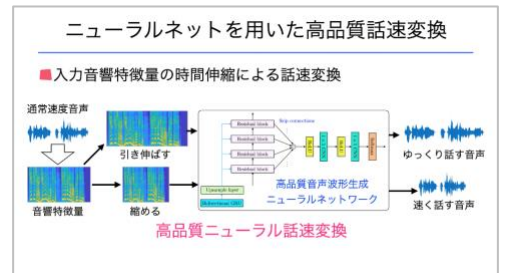
WEB展示では実際の通常速度の音声を、1.25倍の高速再生と0.66倍のスロー再生音声を紹介し、従来方式よりも高品質な話速変換が実現できる実験結果を示し、評価者による音響評価結果において好結果が得られたことを示しました。

また、通常よりもゆっくりと話すことは、多言語音声翻訳とも親和性があります。そのため、この技術は、様々なグローバルコミュニケーションにおいても役立つことが期待されています。

話速変換技術と課題

- 話速変換技術とは
 - 音の高さを変えずに音声を伸縮する技術
- 話速変換技術の必要性
 - 高齢者等はゆっくり話す音声を好む
 - 視覚障がい者等は高速再生の音声を好む
- 課題
 - 従来の方式は時間伸縮するだけでも音質が劣化

高品質ニューラル音声合成を応用した方式を提案



(令和2年11月作成)

問い合わせ先

けいはんな R&D フェア実行委員会 事務局 (NICT 内)

Tel: 0774-98-6900

Email: khn-fair2020@khn.nict.go.jp

https://khn-fair.nict.go.jp/

